

Phil Spector

Data Manipulation with R

 Springer

Contents

Preface	V
1 Data in R	1
1.1 Modes and Classes	1
1.2 Data Storage in R	2
1.3 Testing for Modes and Classes	7
1.4 Structure of R Objects	7
1.5 Conversion of Objects	8
1.6 Missing Values	10
1.7 Working with Missing Values	10
2 Reading and Writing Data	13
2.1 Reading Vectors and Matrices	13
2.2 Data Frames: <code>read.table</code>	15
2.3 Comma- and Tab-Delimited Input Files	17
2.4 Fixed-Width Input Files	17
2.5 Extracting Data from R Objects	18
2.6 Connections	23
2.7 Reading Large Data Files	25
2.8 Generating Data	27
2.8.1 Sequences	27
2.8.2 Random Numbers	29
2.9 Permutations	30
2.9.1 Random Permutations	30
2.9.2 Enumerating All Permutations	30
2.10 Working with Sequences	31
2.11 Spreadsheets	33
2.11.1 The <code>RODBC</code> Package on Windows	33
2.11.2 The <code>gdata</code> Package (All Platforms)	34
2.12 Saving and Loading R Data Objects	35
2.13 Working with Binary Files	36

2.14	Writing R Objects to Files in ASCII Format	38
2.14.1	The <code>write</code> Function	38
2.14.2	The <code>write.table</code> function	39
2.15	Reading Data from Other Programs	39
3	R and Databases	43
3.1	A Brief Guide to SQL	43
3.1.1	Navigation Commands	43
3.1.2	Basics of SQL	44
3.1.3	Aggregation	45
3.1.4	Joining Two Databases	46
3.1.5	Subqueries	47
3.1.6	Modifying Database Records	48
3.2	ODBC	49
3.3	Using the RODBC Package	50
3.4	The DBI Package	51
3.5	Accessing a MySQL Database	51
3.6	Performing Queries	52
3.7	Normalized Tables	52
3.8	Getting Data into MySQL	53
3.9	More Complex Aggregations	55
4	Dates	57
4.1	<code>as.Date</code>	57
4.2	The <code>chron</code> Package	59
4.3	POSIX Classes	60
4.4	Working with Dates	63
4.5	Time Intervals	64
4.6	Time Sequences	65
5	Factors	67
5.1	Using Factors	67
5.2	Numeric Factors	70
5.3	Manipulating Factors	70
5.4	Creating Factors from Continuous Variables	72
5.5	Factors Based on Dates and Times	73
5.6	Interactions	74
6	Subscripting	75
6.1	Basics of Subscripting	75
6.2	Numeric Subscripts	75
6.3	Character Subscripts	75
6.4	Logical Subscripts	76
6.5	Subscripting Matrices and Arrays	77
6.6	Specialized Functions for Matrices	81

6.7	Lists	82
6.8	Subscripting Data Frames	83
7	Character Manipulation	87
7.1	Basics of Character Data	87
7.2	Displaying and Concatenating Character Strings	87
7.3	Working with Parts of Character Values	89
7.4	Regular Expressions in R	90
7.5	Basics of Regular Expressions	91
7.6	Breaking Apart Character Values	93
7.7	Using Regular Expressions in R	94
7.8	Substitutions and Tagging	98
8	Data Aggregation	101
8.1	<code>table</code>	101
8.2	Road Map for Aggregation	106
8.3	Mapping a Function to a Vector or List	107
8.4	Mapping a function to a matrix or array	110
8.5	Mapping a Function Based on Groups	113
8.6	The <code>reshape</code> Package	120
8.7	Loops in R	126
9	Reshaping Data	131
9.1	Modifying Data Frame Variables	131
9.2	Recoding Variables	132
9.3	The <code>recode</code> Function	134
9.4	Reshaping Data Frames	135
9.5	The <code>reshape</code> Package	140
9.6	Combining Data Frames	142
9.7	Under the Hood of <code>merge</code>	146
	Index	149